Princeton University
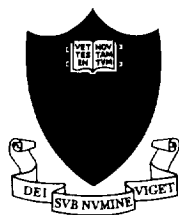
FINAL REPORT: NASA Grant NAG1-1698

STOCHASTIC PREDICTION OF VIBRATION LEVELS IN THE PRESENCE OF MODELING UNCERTAINTIES

Minh Q. Phan, Principal Investigator

Department of Mechanical and Aerospace Engineering

Princeton University

Princeton, New Jersey 08544

November 5, 1997

Department of Mechanical and Aerospace Engineering

FINAL REPORT: NASA Grant NAG1-1698

STOCHASTIC PREDICTION OF VIBRATION LEVELS IN THE PRESENCE OF MODELING UNCERTAINTIES

Minh Q. Phan, Principal Investigator
Department of Mechanical and Aerospace Engineering
Princeton University
Princeton, New Jersey 08544

November 5, 1997

STOCHASTIC PREDICTION OF VIBRATION LEVELS IN THE PRESENCE OF MODELING UNCERTAINTIES

Principal Investigator: Minh Q. Phan

Department of Mechanical and Aerospace Engineering

Princeton University, Princeton, NJ 08544

Abstract

In the design of controllers to suppress unwanted vibration in a flexible spacecraft structure, a nominal model of the system is assumed. Since this design model can only represent the real structure up to a certain level of accuracy, any response predicted by the model is subject to errors that need to be quantified. One such characterization is the probability of the system failing to meet a certain level of performance when the design model is subject to stochastic uncertainties described by certain statistics. Since the direct relationship between the (stochastic) uncertainties in the design model and the resultant responses is very complicated, analytical prediction approaches tend to be limited because many simplifying assumptions must be made to make the problem mathematically tractable. Stochastic robustness analysis via Monte Carlo evaluations can be used to address such a difficulty by converting the stochastic prediction problem into a statistical sampling problem where basic and well-tested principles of probability theory apply. The empirical nature of the procedure eliminates the need for simplifying assumptions. At the same time, however, the empirical results can be shown to obey certain general analytical relationships which allow precise interpretation of the results. This research examines the problem of characterizing the expected vibration level in a flexible structure under periodic disturbances when the parameters in the model are subject to stochastic uncertainties. Analytical expressions describing the statistics of the stochastic estimation is derived. Practical results such as the probabilies of the actual vibration amplitudes failing to meet various threshold levels are obtained together with their associated error bounds. An example will be used to illustrate the application of such a stochastic prediction procedure.

1

Introduction

Periodic disturbance is a major cause of vibrations in a flexible structure. On a spacecraft structure, for example, the disturbance is caused by a particular on-board device performing a periodic scanning motion. The vibrations will be felt by other scientific instruments mounted on the same structure. At the design stage, it is important to quantify the levels of such vibrations in order to determine whether or not there is a need for active control to suppress such periodic disturbances. Typically, only a design (analytical) model of the system is available whose parameters represent those of the real structure up to only a certain level of accuracy. Any response predicted by the model is thus subject to errors that need to be quantified. Should this analysis reveal the need for active control, it is necessary to assess the performance of a particular controller design in the presence of uncertainty in the design model itself. In one case one may be concerned with stability robustness to make sure that the system remains stable in the presence of modeling errors. In another case, performance is a critical issue and one is concerned with the ability of the controller to maintain a certain specified performance goal. These robustness issues are particularly critical for model-based vibration suppression controllers where errors in the design model may result in undesirable effects such as actual amplification rather than attenuation of the periodic disturbances or at worst instability itself. Again, to address these issues adequately, it is important to be able to quantify as precisely as possible how the system stability or performance is affected by the presence of uncertainties in the design model.

This research considers the case where the flexible structure is affected by a persistent disturbance source. The objective the stochastic estimation problem is to determine the expected vibration levels when the parameters in the model are not known accurately but their statistics can be reasonably assumed. For simplicity, the disturbance is

assumed to be known but the approach presented here is equally applicable to the case where uncertainty in the disturbance is present. The ability to estimate the system vibration given uncertainties that are not necessarily specified in the design of a particular controller will provide insights as to how it is expected to function in practice. This ability helps one to evaluate or eliminate particular controller designs.

Analytical solutions to a stochastic estimation problem such as the one considered here are quite difficult. The process tends to require stringent assumptions on the dynamic process and simplifying assumptions must be made to make the mathematics relatively easy to handle. Sensitivity analysis can be used to predict the probability of instability as shown in Refs. 1-4. For a linear system, sensitivity of the closed-loop eigenvalues is used as a measure of (in)stability. The same problem can be approached numerically (or statistically), which involves random sampling of the uncertain parameters according to their specified distributions and then obtaining the statistics of the interested quantities from the sampled results. This is known as Monte Carlo method, which is often used in safety assessment or reliability engineering, Ref. 5-7. In contrast to the analytical approach, the numerical method is flexible in that the distributions of the uncertain parameters need not be assumed to be Gaussian or uncorrelated, the system need not be linear, and any measure of stability or performance can be used. In return for this flexibility, computational expense is incurred. Recently, this method, known as stochastic robustness analysis, has been adopted for use to evaluate stability and performance robustness of control systems and shown to be within the scope of modern computers, Refs. 8-11. This research considers its application to the problem of characterizing structural vibration under periodic disturbance in the presence of stochastic uncertainties in the design model. The model can be either a closed-loop model or an open-loop model depending whether a feedback controller is present. For a flexible spacecraft structure, the multi-mode model is typically derived from a finite element or system identification method. The Monte Carlo-based

prediction approach relies on a fundamental principle of probability that given a sufficiently large number of samples, it is possible to deduce the statistics of the underlying process. It is not necessary to make an exhaustive search before some definite statements about the statistics of the underlying process can be made, although the estimation will become increasingly more accurate as more samples are taken. Furthermore, for processes such as those satisfying a binomial distribution such as the case considered here, it can be shown that the overall statistics obey certain general rules without regard to the number of uncertain factors influencing the outcome of each sample in the first place. In this research, a Gaussian function is used to approximate the binomial distribution so that analytical relations describing the statistics of the underlying process can be explicitly derived. These analytical results allow one to move beyond obtaining merely an estimated probability from a set of samples, but also rigorous confidence bounds on the *actual* underlying probability of the process. This is a very important feature of a Monte-Carlo based estimation approach that is often under-appreciated because at times it may appear counter-intuitive. Furthermore, the effects of any specified parameter variations on any response variable of interest can be directly analyzed for any the disturbance-to-response model. The method returns a set of "user-friendly" measures of the probability of the vibration exceeding various levels, and expected response envelope. The associated probability density function and confidence intervals for other variables of interest, such as the system natural frequencies and damping factors, can be similarly obtained. Numerical examples will be presented to illustrate the applicability of such results.

Finally, it is worth mentioning that statistical prediction has been routinely used in many other fields other than engineering risk and reliablity assessment. One example is the prediction who is expected to win in a presidential election based on exit poll which constitutes a tiny fraction of the voting population. Exit polls, although not perfect, tend to be quite accurate yet the number of factors that influence a person's decision is quite

4

numerous. Another example is the estimation of the percentage of the population that survives by the age of 65. There are countless factors that affect the life span of each individual yet the resultant statistics from a sampled population is sufficiently good to support the typically healthy life insurance industry.

Analysis of Stochastic Vibrations by Monte Carlo Evaluation

Consider a linear time invariant system of the form

$$\frac{dx}{dt} = A(\theta)x(t) + B_1(\theta)u(t) + B_2(\theta)v(t)$$
$$y(t) = C(\theta)x(t)$$

(1)

where $x(t)$ denotes the state vector, $y(t)$ the output vector, $u(t)$ the control input, and $v(t)$ the disturbance input to the system. The system matrices are denoted by $A(\theta)$, $B_1(\theta)$, $B_2(\theta)$, and $C(\theta)$ which are continuous functions of the uncertain parameters in the vector θ. The control input influences the system via a state feedback controller, say

$$u(t) = Gx(t)$$

(2)

The closed-loop system is then governed by the set of equations

$$\frac{dx}{dt} = \left(A(\theta) + B_1(\theta)G\right)x(t) + B_2(\theta)v(t)$$
$$y(t) = C(\theta)x(t)$$

(3)

For generality we have included both control input and disturbance input in the above model. If there is no control and we are only interested in the system open-loop response subject to uncertainties in the design model, then $G = 0$. It is assumed that the implicit

relationship between the uncertain parameters vector θ and the rest of the model is known from a computational viewpoint, i.e., given a set of uncertain parameters it is possible to compute the corresponding system model $A(\theta)$, $B_1(\theta)$, $B_2(\theta)$, and $C(\theta)$. For ease of presentation, the disturbance $v(t)$ is also assumed to be known although uncertainties in the disturbance itself can be treated similarly. The uncertain parameters in θ are not known precisely except for their statistics specified by their probability density functions $pdf(\theta_i)$ for $i = 1,2,...$ up to the number of uncertain parameters. In the spacecraft vibration problem, one wishes to predict the levels of stochastic jitter caused by such disturbances in the presence of the specified stochastic parameter uncertainties. For simplicity of explanation, the magnitude of the steady state response is used as a variable of interest although more sophisticated measures of vibration can be used. Each Monte Carlo run is initiated based on a random generation of the uncertain parameters satisfying the prescribed statistics. To simplify the discussion further, only the maximum amplitude $y = |y_{max}|$ of the each steady state response corresponding to each run is saved for statistical analysis.

Estimation of the statistics of the Monte Carlo evaluations:

The first and most obvious step in the analysis is an examination of the statistics of the results obtained from n Monte Carlo runs, which for the current consideration, consists of n values of $|y_{max}|$. Since the uncertain parameters in the system are continuous random variables, the resultant maximum response is a continuous random variable whose statistics can be described by its probability density function. The difficulty in constructing the density function of a continuous random variable using a finite amount of data (n points in this case) is that the data appear to be discrete. To resolve this problem, the continuous random variable is first converted into a discrete one by segmenting the response into non-overlapping equal intervals Δy. The fraction of measurements in y falling in each respective interval is simply the ratio

6

$$\frac{N(y_{i-1} < y < y_i)}{n}$$

where $N(y_{i-1} < y < y_i)$ denotes the number of times the maximum amplitude of the response fall between y_{i-1} and $y_i = y_{i-1} + \Delta y$, $i = 0,1,2,\dots$ To approximate the probability density function $\hat{p}(y)$ of the original continuous random variable $y = |y_{max}|$, a histogram can be constructed such that the area of each rectangular column in the histogram records the fraction of measurements falling in each respective interval. Thus an estimated probability density function is computed from

$$\hat{p}(y_{i-1} < y < y_i) = \frac{N(y_{i-1} < y < y_i)}{n(\Delta y)} \tag{4}$$

Here the notation $\hat{p}(.)$ is used to denote an estimated probability for the truly correct underlying probability $p(.)$. By examining the shape of the approximated density function of the maximum response, one has a general picture of overall statistics of the responses collected from n Monte Carlo runs. From $\hat{p}(y_{i-1} < y < y_i)$, the corresponding cumulative distribution function $\hat{p}(y < y_i)$ can be computed as

$$\hat{p}(y < y_i) = \sum_{\tau=1}^{i} (\Delta y)\hat{p}(y_{\tau-1} < y < y_\tau) \tag{5}$$

It is known that as the number of data points tends to infinity, the law of large number applies, which will result in the estimated distribution converging to the limiting distribution. In practice it is impossible to have an infinite number of measurements. In the following sections, however, it will be shown that it is possible to quantify precisely how fast one can expect to have the approximated distribution converge to the limiting distribution as a function of the number of Monte Carlo runs.

7

Estimation of the probability of the response exceeding a certain limit:

A reasonable (and practical) question to ask is the probability of the response exceeding a certain specified level, beyond which is considered unacceptable. An estimation of the probability of the maximum response y exceeding a certain level y_i, denoted by $\hat{p}(y \geq y_i)$, is computed from

$$\hat{p}(y \geq y_i) = \frac{N(y \geq y_i)}{n} \tag{6}$$

where n denotes the total number of Monte Carlo runs and $N(y \geq y_i)$ denotes the number of runs where the limit y_i is exceeded. Note that this information is already available from the cumulative distribution estimated in the previous section as

$$\hat{p}(y \geq y_i) = 1 - \hat{p}(y < y_i) \tag{7}$$

which provides the (estimated) probability of the response exceeding not just a single limit but for any limit y_i.

In the previous section, we estimated the statistics of the maximum response amplitude where $y = |y_{max}|$ is treated as a random variable. This is only the first step in the analysis. For each value y_i there exists a truly correct (unknown) underlying probability $p(y \geq y_i)$ which exists independently of any number of Monte Carlo runs made. We can estimate this value of $p(y \geq y_i)$ by Eq. (6) or Eq. (7) increasingly accurately if we could make infinitely many Monte Carlo runs. In practice since only a *finite* number of runs can be made, our estimated $\hat{p}(y \geq y_i)$ is only approximately true. The next step in the analysis is to examine the statistics of $\hat{p}(y \geq y_i)$ itself where $\hat{p}(y \geq y_i)$ is now considered as a random variable.

Estimation of the distribution for $\hat{p}(y \geq y_i)$

Recall that there are n independent runs where n values of $|y_{max}|$ are recorded. Out of these n values, the number of runs whose responses exceed the limit is $N(y \geq y_i)$ which can now be considered a random variable because if we are to conduct several n Monte Carlo runs then $N(y \geq y_i)$ will be different each time. The possible values of $N(y \geq y_i)$ are $0,1,2,\ldots, n$, and as mentioned previously, the ratio for $\hat{p}$ given in Eq. (6) or (7) gives the estimated probability of exceeding the limit y_i. If n is large then $\hat{p}$ will be close to the true probability of exceeding the limit p. When n is not sufficiently large, the estimation $\hat{p}$ will fluctuate. It is therefore important to find the distribution of $\hat{p}$.

First, note that the values of $\hat{p}$ are fractional, but values of $N(y \geq y_i)$ are integers. Also, the probability of $\hat{p}$ taking the value say k/n is the same as the probability of $N(y \geq y_i)$ taking the value k. Therefore it is more convenient to find the distribution of $\hat{p}$ from the distribution of $N(y \geq y_i)$, which specifies the probability of getting $0,1,2, \ldots$ up to (at most) n limit-exceeding responses in n runs. One way to obtain this distribution is by repeating the n independent runs m times where m is a sufficiently large number. It is both unrealistic and necessary to do so. Recognizing that each run is a Bernoulli trial (a Bernoulli trial is one that has only two possible outcomes, namely, the result of each run is that either the maximum response exceeds the limit or it does not exceed the limit), the probability for $N(y \geq y_i)$ to take a particular value k follows a (discrete) binomial distribution ($N(y \geq y_i)$ is a discrete random variable),

$$p[N(y \geq y_i) = k] = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0,1,\ldots,n \tag{8}$$

where $\binom{n}{k}$ denotes the binomial coefficient,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{9}$$

Also, recall that probability of $N(y \geq y_i)$ taking the value k is the same as the probability of $\hat{p}$ taking the value say k / n, i.e.,

$$p[N(y \geq y_i) = k] = p[\hat{p}(y \geq y_i) = k / n] \tag{10}$$

The computation of the binomial coefficient is quite tedious. Numerical overflows occurs for $n > 170$ on a typical computer. Fortunately, it is known the binomial distribution can be approximated by a Gaussian function which is briefly described below. In this research, we will use this approximation in the subsequent analysis.

Approximation of Binomial Distribution by a Gaussian Function

For n larger than 50 the binomial distribution can be closely approximated by a Gaussian function $f_{X,\sigma}(k)$

$$\begin{aligned} p[N(y \geq y_i) = k] &= \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0,1,\ldots,n \\ &\approx \frac{1}{\sigma_X \sqrt{2\pi}} e^{-(k-X)^2/2\sigma_X^2} \equiv f_{X,\sigma_X}(k) \end{aligned} \tag{11}$$

with mean X and variance σ_X,

$$X = np \quad , \quad \sigma_X = \sqrt{np(1-p)} \tag{12}$$

The Gaussian approximation is symmetric about p, implying that the mean value is also the most probable value. Unlike its Gaussian approximation, the binomial distribution is only

10

symmetric for $p = 0.5$. For n larger than 50, however, this non-symmetry is no longer pronounced, and for all practical purposes, the binomial distribution can then be approximated by a Gaussian approximation. In our particular application, n is the number of Monte Carlo runs which is typically several orders of magnitudes larger than 50, the Gaussian approximation $f_{X,\sigma}(k)$ is quite adequate.

Note that the original binomial distribution is discrete. For each value of k, the height of the distribution is $p[N(y \geq y_i) = k]$ records the actual probability of getting k limit-exceeding cases in n Monte Carlo runs. The Gaussian approximation, however, is a continuous function which treats k as a continuous random variable while it is in fact a discrete one. Certain conceptual difficulties may arise due to a fundamental difference between discrete and continuous variables. Namely, for a discrete random variable, $p[N(y \geq y_i) = k]$ is not necessarily zero for some k but for a continuous random variable, $p[N(y \geq y_i) = k]$ must be zero for all k. For this reason care must be taken to interpret various operations involving the continuous Gaussian approximation of a discrete distribution. Specifically, there are two applications of this approximation which deserve special clarification to avoid possible confusion: First, the Gaussian function can be used to compute the probability of getting a certain number of k limit-exceeding cases in n Monte Carlo runs. In this case, the Gaussian function should be viewed only as a numerical approximation of the binomial coefficient, not a continuous distribution. Second, because the increment in the k-axis is 1, the Gaussian function can also be viewed *as if* it is a continuous distribution of a continuous random variable k for the purpose of computing various probabilities. The following example illustrates this point. The probabilities of getting between k_1 and k_2 limit-exceeding cases can be correctly obtained by summing probabilities of getting between k_1 and k_2 limit-exceeding cases specified by the discrete binomial distribution. Since the computing the binomial coefficients is numerically inconvenient for large n, this answer can be approximated by computing the area under the

11

Gaussian curve between k_1 and k_2 treating k as a continuous random variable. This is possible because the integral can be approximated by summing up the appropriate rectangular areas under the curve (using the rectangular approximation to the integral). Since each of the rectangle has a width of 1 unit (because each increment in the k-axis is 1), the area under the curve is numerically the same as the sum of the heights of the corresponding rectangles. Since each height measures the probability of getting k-limit exceeding cases in n runs, this integral approximates the actual probability of getting between k_1 and k_2 limit-exceeding cases as specified by the original discrete binomial distribution. Furthermore, the approximation will become quite good for sufficiently large n as it is the case in our application. We will take advantage of this observation to use the Gaussian approximation to obtain various expressions for the confidence intervals regarding the statistics of the probabilistic estimates. This task is shown in the following sections.

Confidence interval for $\hat{p}$

One key observation of the binomial distribution governing the statistics of the stochastic sampling process is that it is completely specified by the underlying probability p and the number of Monte Carlo runs n. Let us first make a few qualitative observations of the distributions and their subsequent implications. More rigorous results will be presented later. Consider three distributions having the same value of $n = 100$ but with different $p = 0.05$, 0.5, and 0.95. Note that the "width" of the distribution is symmetrical with respect to the axis $p = 0.5$. For $0 < p < 0.5$, this width is smaller as p approaches 0. The same applies for $0.5 < p < 1$ as p approaches 1. Consider a particular case, say $p = 0.5$. The distribution is centered at the true underlying probability $p = 0.5$. For an underlying probability of 50%, one can expect with a very high degree of confidence that with 100 Monte Carlo runs the estimated probability of the system response exceeding a certain limit falls between 0.3 and 0.7. Equivalently, it is unlikely that the statistics obtained from 100

12

Monte Carlo runs will yield a result outside this range if the true underlying probability is 50%. One question that immediately arises is: will the width of the distribution narrows if the number of Monte Carlo runs is increased? The answer is definitely yes. With increasing number of Monte Carlo runs, the width of the distribution becomes narrower, implying that the estimated probability of exceeding a certain limit is even more likely to be close to the true underlying probability. In the following, we quantify these observations by providing analytical expressions for them using the notion of a confidence interval.

Let us consider the case where the underlying probability of exceeding a certain limit y_i is $p(y \geq y_i)$ or p for short, and the number of Monte Carlo runs is n. As established earlier, p and n completely define the distribution of getting a certain number of limit-exceeding cases denoted by $N(y \geq y_i)$ in n runs. This distribution can be approximated by a Gaussian function $f_{X,\sigma_X}(k)$ with the mean X and standard deviation σ_X. For a Gaussian distribution, the standard deviation σ_X corresponds to a 68% confidence level, i.e., there is a 68% probability that the number of limit-exceeding responses $N(y \geq y_i)$ falling between $X - \sigma_X$ and $X + \sigma_X$

$$X - \sigma_X \leq N(y \geq y_i) \leq X + \sigma_X \tag{13}$$

This confidence level is the probability denoted by P and given by the expression

$$
\begin{aligned}
P(X - \sigma_X \leq N \leq X + \sigma_X) &= \frac{1}{\sigma_X \sqrt{2\pi}} \int_{X - \sigma_X}^{X - \sigma_X} e^{-(k-X)^2/2\sigma_X^2} dk \\
&= \frac{1}{\sqrt{2\pi}} \int_{-1}^{1} e^{-z^2/2} dz
\end{aligned}
\tag{14}
$$

where N is used as a short hand notation for $N(y \geq y_i)$. The simplification is obtained by a change of variable from λ to z, $(\lambda - X)/\sigma_X = z$, which simplifies the problem

considerably by making the expression independent of σ_X and X, which in turns depend on n and p. Equation (14) can be generalized for any specified value of confidence level for $N(y \geq y_i)$ falling between $X - \alpha_X$ and $X + \alpha_X$ for any value of $\alpha_X = t\sigma_X$ as

$$P(X - \alpha_X \leq N \leq X + \alpha_X) = \frac{1}{\sqrt{2\pi}} \int_{-t}^{t} e^{-\lambda^2/2} d\lambda \equiv erf(t) \tag{15}$$

The error function or $P = erf(t)$, or its inverse $t = erf^{-1}(P)$, is commonly tabulated in standard statistics.

To convert the above results obtained for N (the number of cases the response exceeds the limit out of n Monte Carlo runs) to $\hat{p}$ (the estimated probability of such occurrence), one only need to divide the results by n. Since,

$$\hat{p} = \frac{N}{n}, \quad p = \frac{X}{n} \tag{16}$$

It follows that $X - \alpha_X \leq N(y \geq y_i) \leq X + \alpha_X$ also implies $p - \alpha \leq \hat{p}(y \geq y_i) \leq p + \alpha$ or

$$P(X - \alpha_X \leq N \leq X + \alpha_X) = P(p - \alpha \leq \hat{p} \leq p + \alpha) \tag{17}$$

provided $\alpha_X = n\alpha$, where we have used the short hand notation $\hat{p}$ for $\hat{p}(y \geq y_i)$. Hence,

$$\begin{aligned}
\alpha &= \frac{\alpha_X}{n} = \frac{t\sigma_X}{n} = \frac{t\sqrt{np(1-p)}}{n} \\
&= t\sqrt{p(1-p)/n} \\
&= erf^{-1}(P)\sqrt{p(1-p)/n}
\end{aligned} \tag{18}$$

14

The difference $(p + \alpha) - (p - \alpha) = 2\alpha$ is the width of the distribution associated with a confidence limit specified by $P = erf(t)$. In short, with n Monte Carlo runs, if the true underlying probability of exceeding a certain limit y_i is p, then there is a $100P\%$ probability that the estimated probability $\hat{p}$ of p will fall between $p - \alpha$ and $p + \alpha$, where $\alpha = erf^{-1}(P)\sqrt{p(1-p)/n}$. Note that the Gaussian function is being used to approximate the underlying binomial distribution and this approximation is taken to be valid for $n \geq 50$.

Given a value of the underlying probability p, the width of the desired confidence interval 2α, and the desired confidence limit P, one can compute the number of Monte Carlo runs required as

$$n = \frac{\left[erf^{-1}(P)\right]^2 p(1-p)}{\alpha^2} \tag{19}$$

Similarly, for each value of the underlying probability p, the width of confidence interval 2α, and the number of Monte Carlo runs n, the expression for the expected confidence level is given by

$$P = erf\left[\frac{\alpha}{\sqrt{p(1-p)/n}}\right] \tag{20}$$

This analysis can be carried out without regard to the number of uncertain parameters in the system. This is a remarkable feature of the discrete binomial distribution under consideration that allows us to make rigorous statements about the nature of the problem and the confidence of the approximation.

15

Confidence Charts

Equation (18) defines the range that an estimated probability $\hat{p}$ must lie within a specified confidence level P given the number of Monte Carlo runs n and the actual underlying probability p, i.e., $p - \alpha \leq \hat{p} \leq p + \alpha$, $\alpha = \alpha(P,p,n)$. Thus for each desired confidence level specified by P, constant P-surfaces of $\alpha = \alpha(P,p,n)$ can be generated. Figure 1 shows two surfaces with confidence levels P equal to 68% and 99.9%. Notice that each constant P-surface is symmetric about the plane $p = 0.50$ and the bound α diminishes rapidly as the number of Monte Carlo runs increases. To view the effect in more details, various projections of the above surface plots can be generated. For particular confidence level P, plots of α versus n corresponding to various values of p can be generated. Figure 2 shows a family of such curves for $p = 50\%$, 10%, 1%, and 0.1% at a confidence level $P = 95.5\%$. Logarithmic scale is used for clarity. In a similar fashion for a particular confidence level, one can also construct plots of α versus p for various values of n. This is shown in Fig. 3 where $n = 100$, 1000, 10000, and 10000 at a confidence level of 95.5%. These charts together provide a general picture of the overall statistics governing the stochastic estimation process that may not be as evident from an equation alone.

Bounds on the Underlying Limit-Exceeding Probability p

The above analysis determines the width of the confidence interval associated with each value of the true underlying probability of exceeding a certain limit. The analysis yields a global picture of the statistics of the estimated result based on a certain number of Monte Carlo runs. Among other things, associated with each value of the underlying probability, it clearly describes how the width of the confidence interval diminishes as the number of Monte Carlo runs increases. In reality, however, the true value of the underlying probability is not known. This fact will introduce further uncertainty in the interpretation of the result. The following example illustrates this point. Let us consider

the following two cases where one evaluates $n = 1000$ Monte Carlo runs and the desired confidence level is 99.7%, i.e., $P = 0.997$. First, consider the case where the true underlying probability $p_1 = 0.05$ or 5%. It can be computed from Eq. (18) that $\alpha_1 = 0.020$ or 2%. Thus, there is a 99.7% probability that an estimated probability of exceeding a certain limit denoted by $\hat{p}$ computed from 1000 Monte Carlo runs will fall between 3% and 7% ($p_1 - \alpha_1 = 0.03$ and $p_1 + \alpha_1 = 0.07$). Equivalently, the probability of $\hat{p}$ falling outside this range is 0.03%. Second, examine the case where the true underlying probability $p_2 = 8\%$. From Eq. (18), $\alpha_2 = 0.025$, which implies that there is *also* a 99.7% probability that an estimated $\hat{p}$ will fall between $p_2 - \alpha_2 = 0.055$ or 5.5% and $p_2 + \alpha_2 = 0.105$ or 10.5%. Let us suppose that with 1000 Monte Carlo runs, we compute $\hat{p}$ and it happen to be 4%. Subject to a 99.7% confidence level, between the two choices it is more likely that the true underlying probability of the process is 5% rather than 8% because 0.04 falls within the range [0.03, 0.07] associated with $p_1 = 0.05$ but not [0.055, 0.105] associated with $p_2 = 0.08$. A more complicated situation will arise if the estimated probability $\hat{p}$ turns out to be 6% which falls within both the ranges [0.03, 0.07] and [0.055, 0.105]. This situation points out the need to carry the analysis one step further to establish the upper and lower bounds on the *true* underlying probability p given a value of the *estimated* probability $\hat{p}$. Again, it is important to emphasize that such a bound on p is based on the estimated probability $\hat{p}$ rather than the true underlying probability p as done previously. The establishment of such bounds is done below.

Suppose a desired confidence level P is given and p is the (unknown) underlying probability of exceeding a certain limit. From n Monte Carlo runs, suppose that $\hat{p}$ is an estimated value of p. For every value of $\hat{p}$ within this confidence level, the underlying probability p must be bounded from below by p_{min} and above by p_{max}, where $p_{min} = \hat{p} - \alpha_{min}$ and $p_{max} = \hat{p} + \alpha_{max}$, i.e.,

$$p_{min} = \hat{p} - \alpha_{min} \le p \le \hat{p} + \alpha_{max} = p_{max} \qquad (21)$$

The lower bound p_{min} and upper bound p_{max} must satisfy the following relationships

$$\alpha_{min} = \hat{p} - p_{min} = erf^{-1}(P)\sqrt{p_{min}(1 - p_{min})/n} \qquad (22)$$

$$\alpha_{max} = p_{max} - \hat{p} = erf^{-1}(P)\sqrt{p_{max}(1 - p_{max})/n} \qquad (23)$$

Making use of these relationship one can solve for the lower bound p_{min} and the upper bound p_{max} in terms of $\hat{p}$, P, and n. To solve for p_{min} from Eq. (22), one can square both sides of the equation and arrange it in the form of a quadratic equation,

$$ap_{min}^2 + bp_{min} + c = 0 \qquad (24)$$

where

$$a = 1 + \frac{\left[erf^{-1}(P)\right]^2}{n} \quad , \quad b = -2\hat{p} - \frac{\left[erf^{-1}(P)\right]^2}{n} \quad , \quad c = \hat{p}^2 \qquad (25)$$

Similarly to solve for p_{max}, one can repeat the same procedure to produce

$$ap_{max}^2 + bp_{max} + c = 0 \qquad (26)$$

with the same coefficients a, b, and c. The implication of this fact is that the solutions for p_{min} and p_{max} are simply

$$p_{min} = \min\left\{\frac{-b - \sqrt{b^2 - 4ac}}{2a}, \frac{-b + \sqrt{b^2 - 4ac}}{2a}\right\} \qquad (27)$$

18

$$p_{max} = \max\left\{\frac{-b-\sqrt{b^2-4ac}}{2a}, \frac{-b+\sqrt{b^2-4ac}}{2a}\right\} \tag{28}$$

These are the bounds on the actual underlying probability p of exceeding a certain limit subject to a confidence level specified by P, i.e., $p_{min} \le p \le p_{max}$. The bounds are computed based on an estimated value of the probability of exceeding the limit. This is a key result.

Thus for any number of Monte Carlo runs conducted and any confidence level desired, the lower and upper limits p_{min} and p_{max} within which the actual underlying probability p must lie become known once an estimated value $\hat{p}$ is obtained according to Eq. (6) or (7). Figure 4 shows the upper and lower limits that bound the actual underlying probabilities at 95.5% confidence level as a function of the estimated probability obtained from 1000 Monte Carlo runs. The fact that the bounds on p are based on $\hat{p}$ and $\hat{p}$ is practically always in error (sometimes substantially so for a "small" finite n) does *not* imply that the bounds are in error. Indeed the above analysis has shown that for a given the number of Monte Carlo runs and the confidence level desired, the bounds on p will be correct irrespective of what $\hat{p}$ turns out to be. For example, with 1000 Monte Carlo runs, suppose that the estimated probability of exceeding a certain limit turns out to be 1%, then it can be stated with 95.5% confidence that the true probability of the process lie between 0.5% and 1.8% . This is a rigorous result in the sense that the statement can be made with 100% certainty (without error) even if the upper and lower bounds are computed from $\hat{p} = 1\%$ which is only an estimated value. The width of this bound ($p_{max} - p_{min}$) can be narrowed at the expense of additional Monte Carlo runs or lowering the confidence level. With 10000 Monte Carlo runs, an estimated probability of 1% implies that the true probability lies between 0.8% and 1.2% with the same confidence level. If one reduces the

confidence level to 68.3% then with the same estimated probability of 10% it can be concluded that the true probability lies between 0.9% and 1.1%.

We have established the bounds on the underlying probability p of exceeding a certain limit based on an estimated probability $\hat{p}$ of exceeding that limit from n Monte Carlo runs. Since the actual bounds on p depends on $\hat{p}$, it is practically certain that a different bound will be obtained if we are to conduct another n Monte Carlo runs. Different bounds obtained from two different sets of n Monte Carlo runs must be consistent in that the true underlying probability p must lie within both bounds. In general, if one carries out n Monte Carlo runs ℓ times then, the bounds can be narrowed further by taking the intersection of the previous bounds, i.e.,

$$p_{min}^{*} = \max\left\{p_{min}^{(1)}, p_{min}^{(2)}, \ldots, p_{min}^{(\ell)}\right\}$$

$$p_{max}^{*} = \min\left\{p_{max}^{(1)}, p_{max}^{(2)}, \ldots, p_{max}^{(\ell)}\right\}$$

Alternatively, one can compute an average limit-exceeding probability from the ℓ sets, i.e.,

$$\hat{p} = \frac{1}{\ell} \sum_{i=1}^{\ell} \hat{p}^{(i)},$$

and then the probability bounds associated with this averaged result for $n\ell$ Monte Carlo runs.

In the above analysis, the number of uncertain parameters clearly did not enter the arguments leading to Eqs. (27) and (28) that establish the lower and upper bounds on the actual underlying probability at a specified confidence level. This rather counter-intuitive assertion can be understood by realizing that the shape of the binomial distribution is

20

completely determined by the number of Monte Carlo runs and the actual (or true) underlying probability regardless of the number of uncertain parameters. The established limits given in Eqs. (27) and (28) merely bound the true probability without implying where it is likely to be placed within these bounds. The number of uncertain parameters does influence the result in the sense that they affect the relative placement of the actual probability within these bounds, but not the bounds themselves. As an illustration, consider the case where 1000 Monte-Carlo runs are conducted from which an estimated probability of 10% is obtained. Subject to a user-specified level of confidence, say 99.9%, it can be said that the true probability is bounded below by 7.3% and above by 13.6%. If the problem has only a few number of uncertain parameters then it is more likely that the estimated probability is close to the true answer, or equivalently the true probability is close to the average of the upper and lower limits, or 10.4%. On the other hand, if the number of uncertain parameters is large, then the same limits still apply but now it is less likely that the true probability is close to the average of the two limits, or 10.4%. In fact, the true probability might very well be closer to the lower limit of 7.3%, or the upper limit of 10.4% because the estimated probability is more likely to be scattered from its true value. This increase or decrease in the degree of uncertainty of the true probability relative to the average of the two limits is the mechanism by which the number of uncertain parameters affect the interpretation of the results. Fortunately, however, it is important to realize that it does not affect the actual bounds on the true probability established in Eqs. (27) and (28).

Further Discussion and Extensions

We have discussed a number of basic elements that will contribute to the prediction of expected vibration levels of a flexible structure in the presence of stochastic modeling uncertainties. The structure is subjected to persistent or periodic disturbance sources. The most critical factor in this analysis is the number of runs needed to gather statistics from which conclusions can be drawn. However, by exploiting the feature of the binomial

distribution governing the uncertainties of the estimates, definite statements about the empirical results can be made rigorously. These types of analysis have been utilized extensively in such diverse applications as opinion surveys and quality control where definite probability conclusions can be drawn based on a small number of samples. Also, in the interest of computation efforts, customized solutions can be devised so that any duplication in the computation can be avoided. For example, for a large dimensional system, if the uncertainties in the design model are limited to a certain number of modes then it is only necessary to compute the contribution of these uncertain modes on the overall solution. Also, in addition to the theoretical results presented here, various indicators can be developed to monitor and verify the convergence of the solution so that it is not necessary to continue the sampling if a point of diminishing return has been reached.

The above analysis provides a prediction on the statistics and associated confidence statements of the expected performance for a particular limit on the response. One can repeat the same analysis on several different limits which are not necessarily the real threshold limits to arrive at a much more general set of results. For example, we can obtain information such as the probability of the vibration amplitude exceeding 1 unit is 50%, but the probability of exceeding 2 units is 10%, and the probability of exceeding 3 units is only 1%, etc... each with each own error bounds. Furthermore, up to this point we have only discussed the maximum amplitudes of the dynamic responses based on a number of runs. In fact, there are more available information from each run than just the maximum amplitude. Statistical analysis and prediction can be carried out to analyze this information which is already available and no additional Monte Carlo runs are needed. Therefore, the additional computation cost to analyze this information is relatively minimal compared with the amount of useful information it returns.

Illustrative Example

To illustrate an application of the stochastic estimation procedure the following example is used. Figure 5 shows a two degree-of-freedom flexible system with eight parameters (two masses, three springs, and three dashpots) all of which are assumed to be uncertain. The system is known to be disturbed by a periodic forcing input at the first mass, $u(t) = \sin(t)$, and we are interested in estimating the amplitude of the steady-state vibration at the second mass. Each uncertain parameter is assumed to follow a (normal) Gaussian distribution, with mean value $\overline{m}_1 = \overline{m}_2 = 1$, $\overline{k}_1 = \overline{k}_2 = \overline{k}_3 = 1$, and $\overline{c}_1 = \overline{c}_2 = \overline{c}_3 = 0.1$. For simplicity, the distribution is taken to be finite and each parameter is allowed to vary up to ±50% around the mean value, i.e., $\Delta m_1 = \Delta m_2 = 0.5$, $\Delta k_1 = \Delta k_2 = \Delta k_3 = 0.5$, $\Delta c_1 = \Delta c_2 = \Delta c_3 = 0.05$. For example, the first mass is allowed to vary according to the rule

$$m_1 = \overline{m}_1 + \left(\frac{\Delta m_1}{4} \right) z$$

where $\overline{m}_1 - \Delta m_1 \leq m_1 \leq \overline{m}_1 + \Delta m_1$, z is a normally distributed random number with mean 0 and standard deviation 1, and the extrema are taken to be four times the standard deviation of the distribution. Figures 6 and 7 show the results obtained with 10^2, 10^3, 10^4, and 10^5 Monte Carlo runs. From each set of Monte Carlo runs, a record of 10^2, 10^3, 10^4, and 10^5 amplitudes of the steady-state responses is recorded for analysis. The left figures show the (estimated) probability density function from the raw data as shown in Eq. (4) from which the corresponding cumulative distribution functions (not shown here) are computed from Eq. (5). These results are then used to compute the estimated limit-exceeding probabilities given by Eq. (7) as shown on the right figures. Based on the estimated limit-exceeding probabilities and the number of Monte Carlo runs, we can compute the maximum and minimum bounds on the *actual* or *true* underlying limit-exceeding probabilities for any

23

level of confidence desired using Eqs. (27) and (28). These bounds are also shown on the right figures of 17-18 for a 99.9% confidence level although they can be computed for any confidence levels just as easily. For example, with 100 Monte Carlo runs, one can estimate the limit-exceeding probabilities as shown in the top right of Fig. 6. Although the estimates are not correct, we can nevertheless establish that there is a 99.9% probability that the true limit-exceeding probabilities will not exceed the estimated probabilities by approximately $\pm 10\%$ (the exact values varies according to $\hat{p}$ according to Eqs. (27) and (28) as plotted in the figure). With more Monte Carlo runs, the bounds can be narrowed down substantially as each successive bounds are nested within the previous bounds. Note that the actual width of the bounds $\left(p_{max} - p_{min}\right)$ depends on the estimated probabilities as shown in Fig. 4. The width is the largest at $\hat{p} = 0.5$ and the smallest at $\hat{p} = 0$ or 1 representing the smallest accuracy level that can be ascertained with the number of Monte Carlo runs evaluated. This measure is comparable to the "resolution" of this estimation method by statistical sampling. With 10^4 Monte Carlo runs, this "resolution" is 0.1%, and with 10^5 runs, it is 0.01%. These limits are computed with a 99.9% confidence level $(P = 0.999)$. If one is willing to reduce the confidence level then this resolution can be narrowed further. For example, in the case of 10^5 Monte Carlo runs, it is 0.003% for a 90% confidence level.

The same statistical estimation procedure can be used to estimate other effects due to the specified variations in the uncertain parameters other than the amplitudes of the steady-state response. For example we may be interested in the distribution of the natural frequencies or the damping factors as a result of the uncertainty in the physical parameters. Such results are shown in Fig. 8 for the estimated frequency distribution and in Fig. 9 for the damping factor distribution of the first mode obtained with 10000 Monte Carlo runs. More importantly, we can also compute the upper and lower limits that bound the actual probabilities underlying the process for any level of confidence desired. These bounds are

24

shown in Figs. 8-9 for a 99.9% confidence level. This kind of information is useful to design robust controllers or to estimate the likelihood that certain periodic disturbance sources may cause the system to vibrate in resonance.

Concluding Remarks

From the point of view of the control system designer, given the statistics of the uncertain parameters, he or she wishes to know the statistics of the expected performance, say in keeping the vibration within a certain range when the system is affected by a disturbance source. Since the relationship between a particular variation in the design model and the response of the system is quite complicated, it is extremely difficult to quantify the affect analytically. What is needed is a flexible approach that can handle these types of question naturally without imposing a lot of unrealistic conditions for the sake of mathematical simplification. In this approach, through the sampling process, the stochastic estimation problem is converted into a probability and statistics problem where there are well-established theoretical results and useful applications. The price to pay for the flexibilities of the method and the type of direct and intuitively appealing answers it provides is computational cost, which should become less expensive over time as faster processors are available. At the present time a typical workstation is adequate for most applications. In this research, we have adapted this statistical approach to estimate the limit-exceeding probabilities of the dynamic response when uncertainties in the analysis model are specified probabilistically. Starting with the estimated probabilities obtained from a number of Monte Carlo runs, it is possible to obtain bounds on the actual underlying probability for any level of confidence desired. By using a Gaussian approximation to the binomial distribution, this research provides analytical expressions for such bounds as well as those describing the precise trade-off between the number of Monte Carlo runs conducted, the accuracy of the estimation, and the level of confidence desired.

Note that the binomial distribution governs the statistics of a Bernoulli variable describing an outcome which can be characterized as a success or a failure. This is precisely the case where we are interested in whether or not the vibration exceed a *particular* level, or the system natural frequency fall within a certain range, etc.... This is to be distinguished from the final distribution of the probabilities exceeding various *different* levels which need not be Gaussian such as the one shown in Fig. 18. Likewise, the statistics of the uncertain variables does not have to be Gaussian, either. A numerical example has been shown to illustrate an application of the stochastic estimation procedure.

The statistical sampling process considered here assumes that the distributions of the uncertain parameters are known or can be assumed before hand. It remains to examine the sensitivities of such an assumption on the final results. Also, the analysis has shown that the resolution of the results depends on the number of Monte Carlo runs conducted. For example, if such a resolution is 0.01%, this sampling method can only establish that the probability of the vibration level exceeding say 10 units *or more* is 0.01%. In practice, if it can be assumed that the distributions of the uncertain parameters are finite, then it is possible to establish the maximum possible amplitude that can occur. This statistical procedure together with an optimization method such as genetic algorithm can further establish a bound on the actual maximum amplitude itself in addition to a bound on the actual probability of such an occurrence as done here. Finally, it is possible to speed up the computation by discretization, i.e., allowing the uncertain parameters to take only distinct values. The considerations mentioned here remain to be examined in further details.

References

1. Morgan, B.S., Jr., "Sensitivity Analysis and Synthesis of Multivariable Systems, " *IEEE Transactions on Automatic Control,* Vol. 11, No. 4, 1966, 506-512.

2. Lim, K.B. and Junkins, J.L., "Probability of Stability: New Measures of Robustness for Linear Dynamical Systems," *Journal of the Astronautical Sciences,* Vol. 35, No. 4, 1987, pp. 383-397.

3. Gilbert, E.G., "Conditions for Minimizing the Norm Sensitivity of Characteristic Roots," *IEEE Transactions on Automatic Control,* Vol. 29, No. 7, 1984, pp.658-661.

4. Pierre, C., "Root Sensitivity to Parameter Uncertainties: A Statistical Approach," *International Journal of Control,* Vol. 49, No. 2, 1989, pp. 521-532.

5. Brown, G.W., "Monte Carlo Methods," *Mondern Mathematics for the Engineer,* edited by Beckenbach, E.F., McGraw-Hill, New York, 1956.

6. Apostolakis, G., "The Concept of Probability in Safety Assessments of Technological Systems," *Science,* Vol. 250, 1990, pp. 1359-1364.

7. Henley, E.J. and Kumamoto, H., *Probabilistic Risk Assesssment: Reliability Engineering, Design, and Analysis,* IEEE Press, New York, 1992.

8. Stengel, R. "Some Effects of Parameter Variations on the Lateral-Directional Stability of Aircraft, " *Journal of Guidance and Control,* Vol. 3, No. 2, April 1980.

9. Ray, L.R and Stengel, R., "Stochastic Measures of Performance Robustness in Aircraft Control Systems," *Journal of Guidance, Control, and Dynamics*, Vol. 15, No. 6, November-December 1992.

10. Stengel, R. and Marrison, C.I., "Robustness Solutions to a Benchmark Control Problem," *Journal of Guidance, Control, and Dynamics*, Vol. 15, No. 5, September-October 1992.

11. Stengel, R., Ray, L.R, and Marrison, C.I., "Probabisitic Evaluation of Control System Robustness," *International Journal of System Science*, Vol. 26, No. 7, 1995.

12. Taylor, J.R., *An Introduction to Error Analysis*, University Science Books, Mill Valley, CA, 1982.

13. Johnson, N.I. and Kotz, S., *Discrete Distributions*, Houghton Mifflin Company, Boston, 1969.
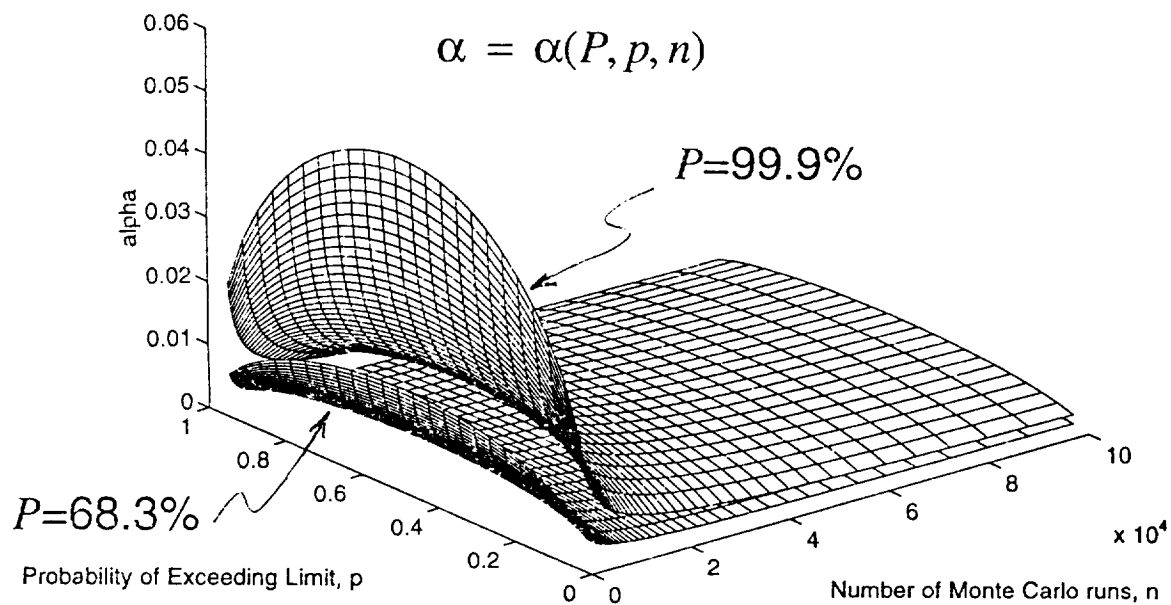
Figure 1: Two constant-confidence surfaces at confidence levels of 99.9% and 68.3%
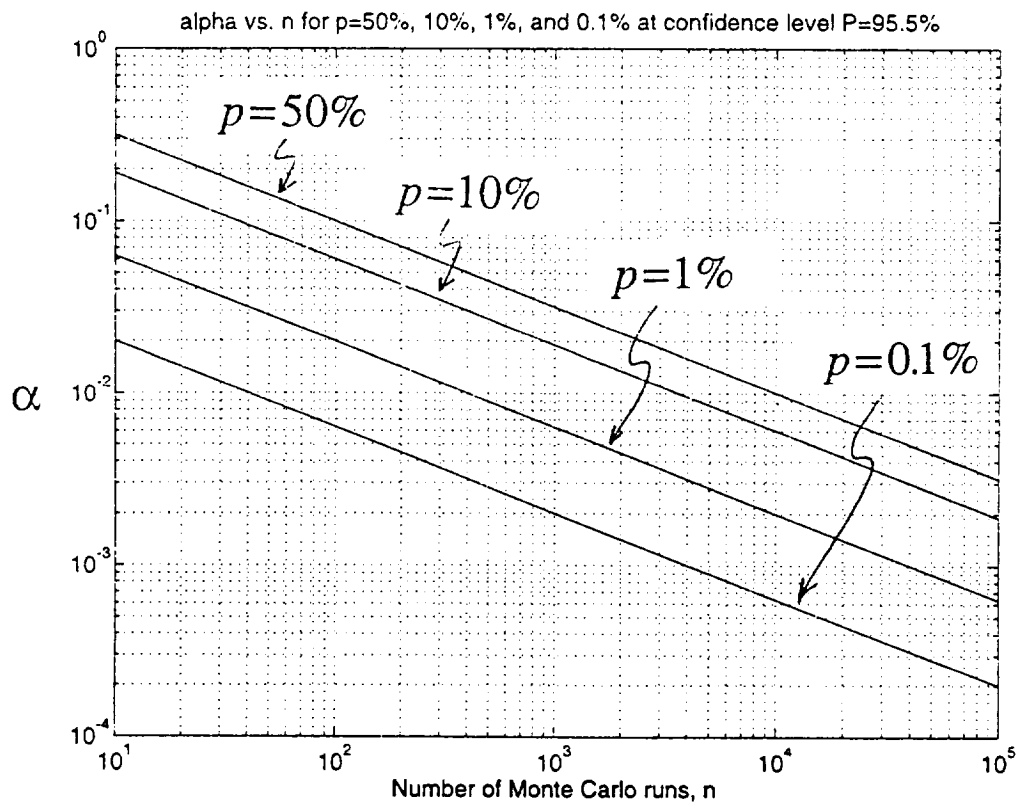
Figure 2: "Width" of distributions versus the number of Monte Carlo runs for various underlying probabilities at 95.5% confidence level.

29
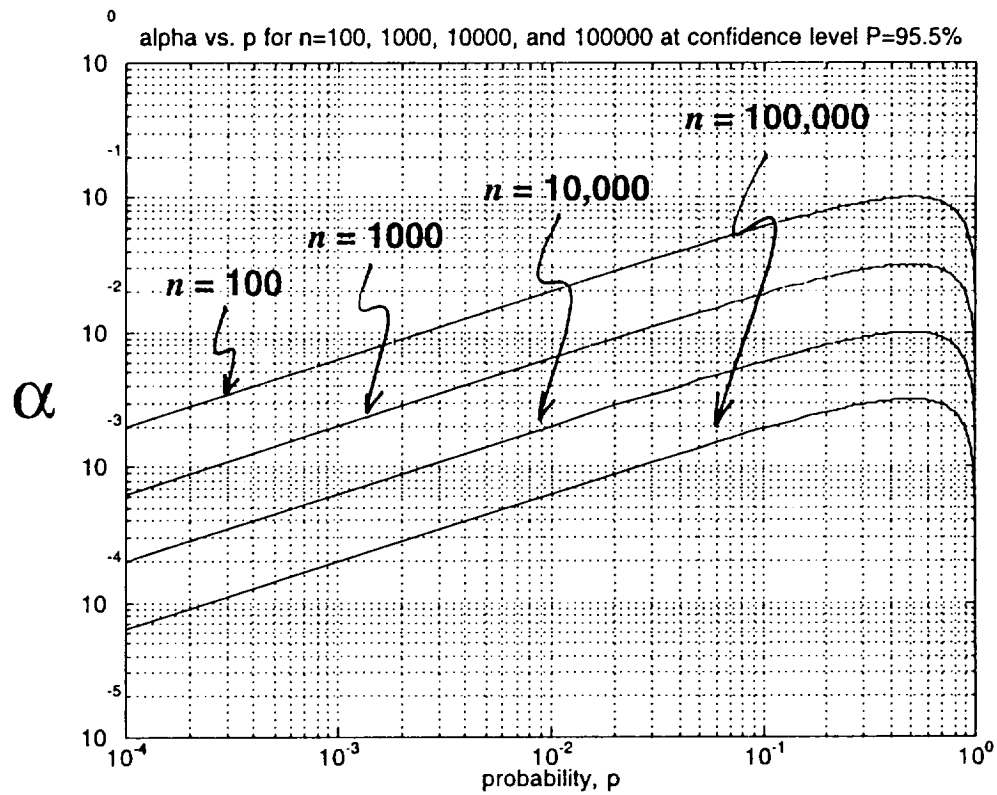
Figure 3: "Width" of distributions versus underlying probability for various numbers of Monte Carlo runs at 95.5% confidence level.
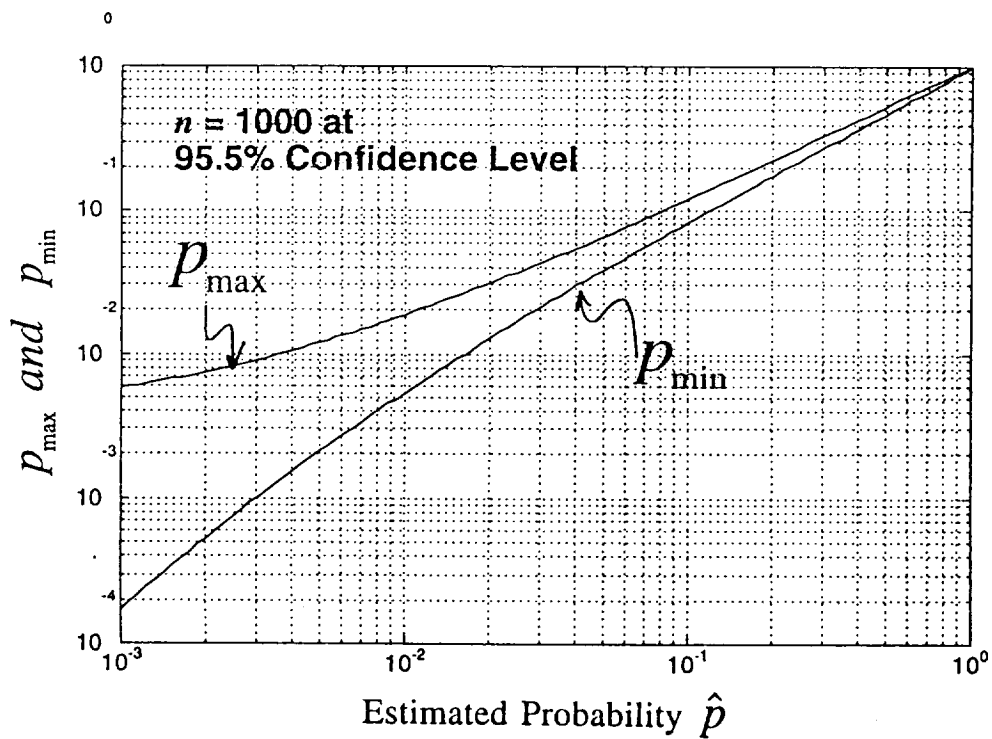

Figure 4: Upper and lower limits (at 95.5% confidence level) that bound the underlying probabilities as function of the estimated probability obtained with 1000 Monte Carlo runs
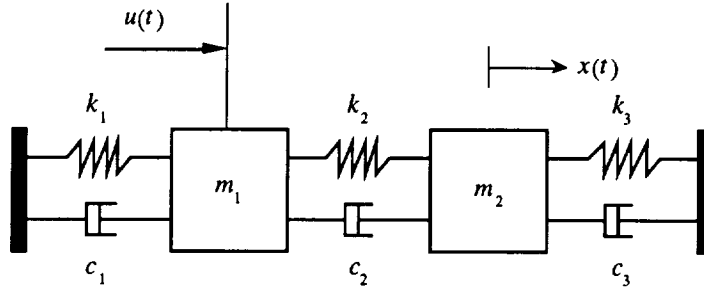
Figure 5: A two-degree-of-freedom system with eight uncertain parameters
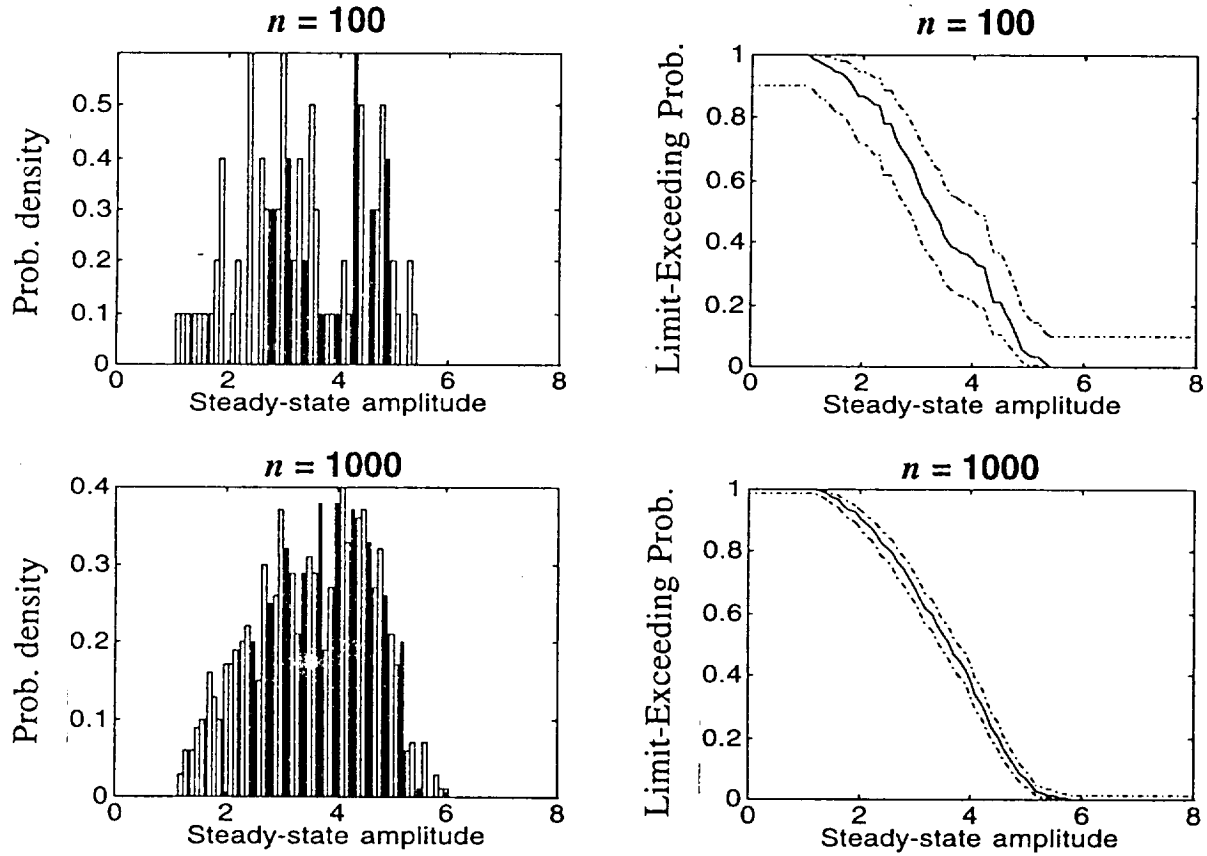
Figure 6: Estimated limit-exceeding probabilities and bounds
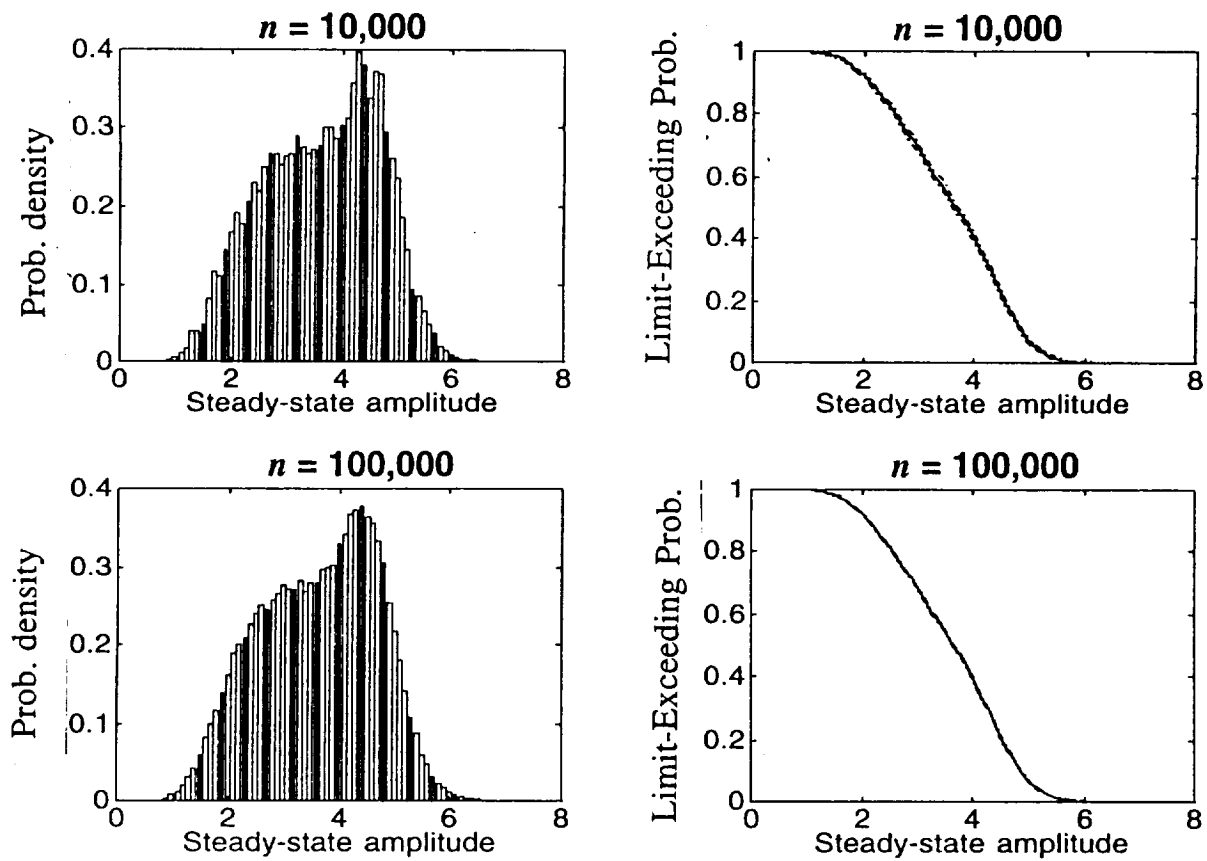at 99.9% confidence level obtained with 100 and 1000 Monte Carlo runs

Figure 7: Estimated limit-exceeding probabilities and bounds
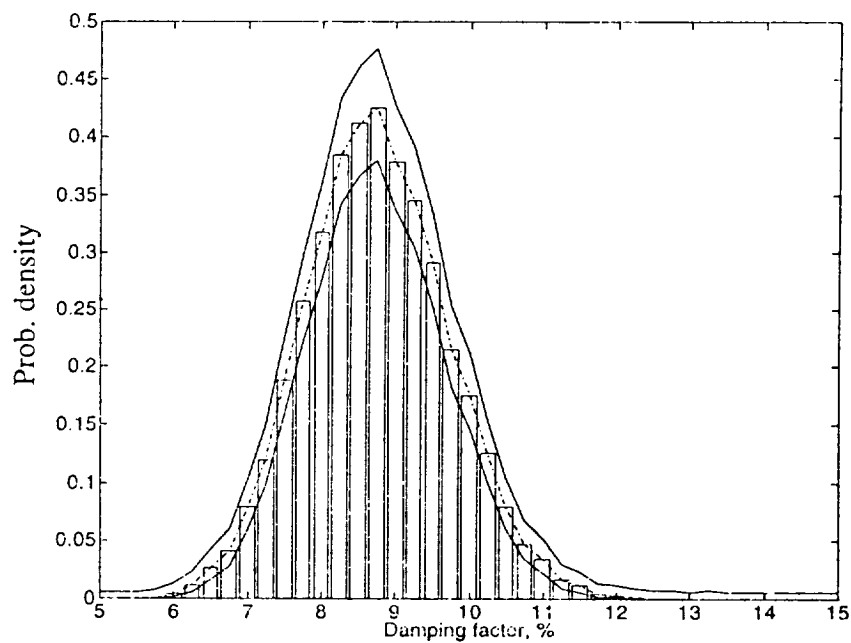at 99.9% confidence level obtained with 10000 and 100000 Monte Carlo runs

Figure 8: Estimated probability density and bounds of the first-mode frequency
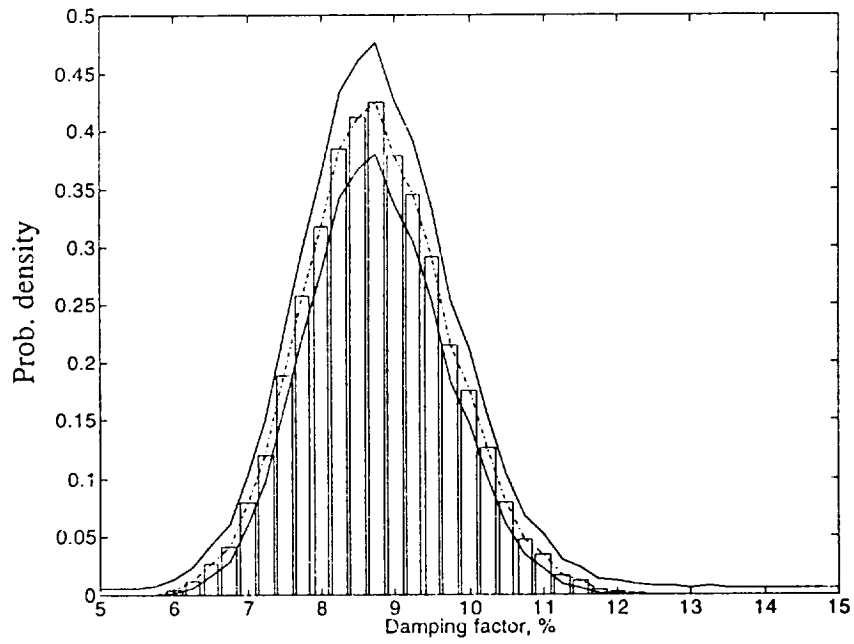at 99.9% confidence level obtained with 1000 Monte Carlo runs

Figure 9: Estimated probability density and bounds of the first-mode damping factor at 99.9% confidence level obtained with 1000 Monte Carlo runs